

Open Access : Compliance Support

Read metadata tags on Publisher PDFs

Released : 5.6.0 / 4.25.0

Document updated : 16 June 2016

Read metadata tags on Publisher PDFs

Publisher PDFs (Proofs (P), Versions of record (VoR), and in some cases Author Accepted Manuscripts (AAM)) are starting to be produced with metadata tags that platforms can use to automatically detect certain OA properties. Where a user is uploading a file with such tags, Pure will populate the Document version and License to document, and derive the Public access to file, as per the properties embedded in the PDF.

This functionality will only be implemented from this point forward – Pure will not go back to read any previously uploaded files.

Publishers

At this stage, it is understood that Elsevier is the only publisher using PDF metadata tags to capture Open Access elements. However other publishers are expected to start making use of such tags from late 2016.

As such, the rules and logic described below currently align with Elsevier's implementation of PDF metadata tags, which adhere to the relevant NISO standards¹. Any PDFs produced by other publishers adopting these same standards should be read automatically, with no new development required by the Pure team. We will continue to monitor this space.

At present, Elsevier are only applying Open Access PDF metadata tags to Proofs and Versions of record.

While Elsevier only intends to use Open Access PDF metadata tags for Journal articles, Pure does not restrict the reading of PDF metadata tags to only Journal articles. Pure will populate the relevant fields for any output type where the uploaded PDF contains the relevant metadata tags.

¹ Elsevier has adopted the NISO RP-22-2015 Access License and Indicators (ALI) standard to include elements to express Open Access information, as well as the NISO RP-8-2008 Journal Article Version (JAV) standard to indicate the stage/phase of journal articles.

Mapping rules

The metadata tags available on the Proofs and VoRs will be used to populate fields in Pure. Unless otherwise stated, the fields will be editable, and not 'locked'. The following rules apply:

File metadata tag		Pure content	
Tag name	value	Field	Pre-select value
jav:journal_article_version	VoR	Document version LOCKED	Final published version
	P		Proof
ali:url		License to document	As we hold the license url in the Document license Description, Pure infers the Document license from the url provided.
			Where the url provided does not match any existing urls, Document license is populated as 'Other' and the URL to license reference with the url provided.
			Where there is no ali:url tag, the Document license remains as 'not set'.

Note that Pure looks to the License URI description to make the match, so you must ensure that the URI descriptions in your classification match that outlined at the end of this document.

Logic

The metadata tags and the values in them are used to determine the accessibility of the file. The following logic is employed:

File metadata tag			Pure content
jav:journal_article_version	ali:url ²		Public access to file
VoR OR P	AND	http://creativecommons.org/licenses/by/4.0/ ... http://creativecommons.org/licenses/by-nd/4.0/ ... http://creativecommons.org/licenses/by-nc/4.0/ ... http://creativecommons.org/licenses/by-nc-nd/4.0/ ... http://creativecommons.org/licenses/by-nc-sa/4.0/ ... http://creativecommons.org/licenses/by-sa/4.0/ ...	THEN Open
VoR OR P	AND	<i>None of the values above</i>	THEN Not set
VoR OR P	AND	<i>tag not present</i>	THEN Closed

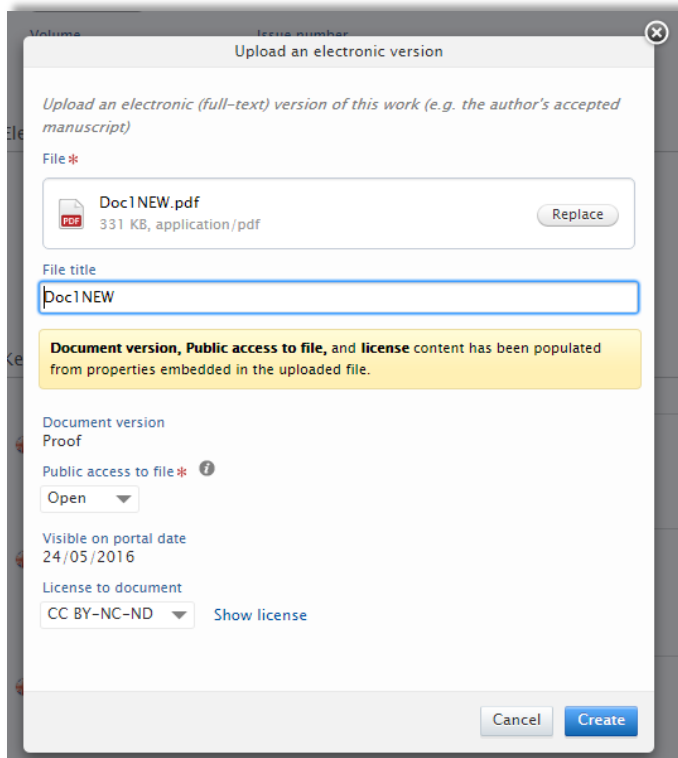
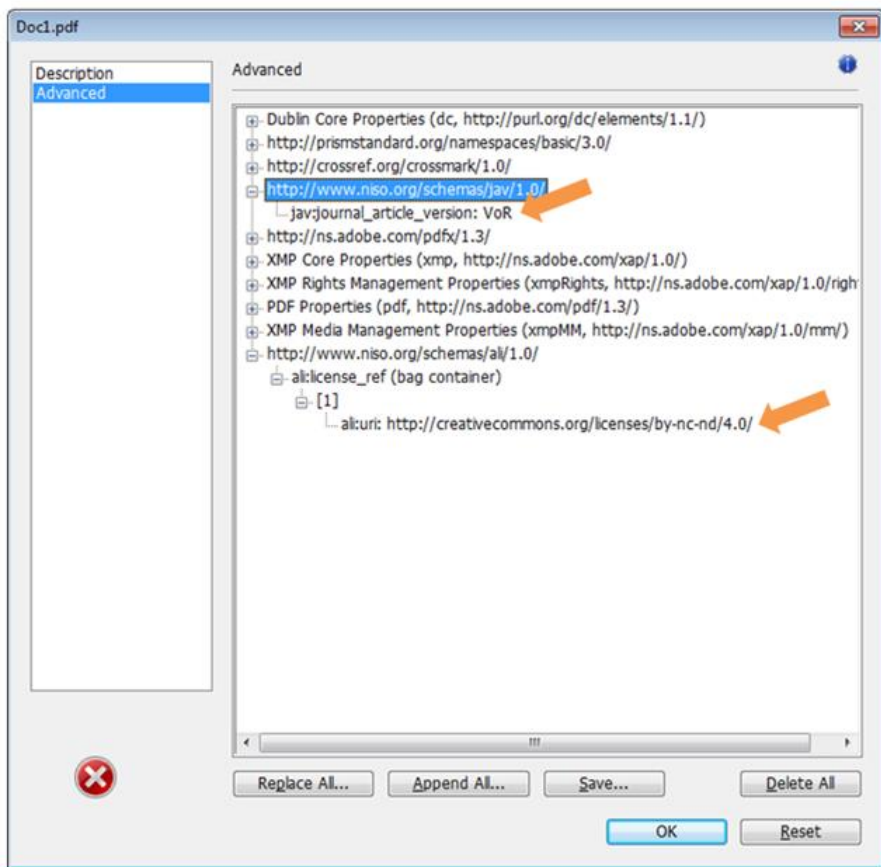
Where Pure has populates fields as per the logic and mapping above, users are advised via an information box on screen.

In the unlikely event where a user has input content (e.g. Document version, License to document) prior to uploading the file, upon file upload the fields will be overwritten with content as per the PDF metadata tags.

Functionality has been updated such that, where you have enabled Coversheets, the original PDF metadata tags will remain embedded in the file upon creation of the coversheet (in previous versions, all tags were deleted).

² Note that the urls cited here are not the full urls. Licenses go through different versions (e.g. the CC BY license is currently on its 4th version (<http://creativecommons.org/licenses/by/4.0/>), and therefore to ensure this functionality continues to operate through new license versions, Pure doesn't look to match the whole url, just the 'trunk'.

Example of PDF metadata tags, as viewed in the PDF Properties > Additional metadata



License URIs

Note that Pure looks to the License URI description to make the License to document match, so you must ensure that the URI descriptions in your classification match the following:

Document Licences
Classification scheme

Scheme

URI *
 /dk/atira/pure/core/document/licenses
The URI must be in a format like /dk/atira/pure/keywords/my_structured_keywords where spaces and capitalised letters are not valid.

Type
 Taxonomy ▼

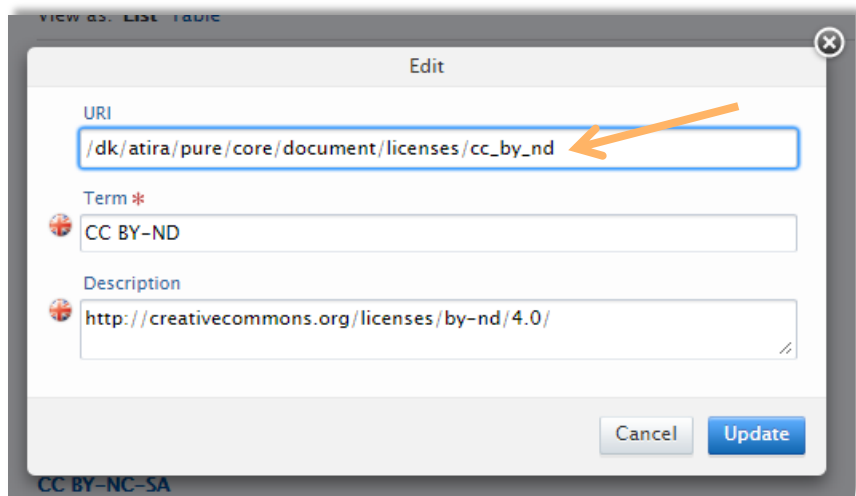
Description
 Document Licences

Classifications

View as: **List** Table

Unspecified <small>/dk/atira/pure/core/document/licenses/unspecified</small>	Add relation... Edit ↓ ↑ —
CC BY-ND <small>http://creativecommons.org/licenses/by-nd/4.0/ /dk/atira/pure/core/document/licenses/cc_by_nd</small>	Add relation... Edit ↓ ↑ —
CC BY <small>http://creativecommons.org/licenses/by/4.0/ /dk/atira/pure/core/document/licenses/cc_by</small>	Add relation... Edit ↓ ↑ —
CC BY-NC <small>http://creativecommons.org/licenses/by-nc/4.0/ /dk/atira/pure/core/document/licenses/cc_by_nc</small>	Add relation... Edit ↓ ↑ —
CC BY-NC-ND <small>http://creativecommons.org/licenses/by-nc-nd/4.0/ /dk/atira/pure/core/document/licenses/cc_by_nc_nd</small>	Add relation... Edit ↓ ↑ —
CC BY-NC-SA <small>http://creativecommons.org/licenses/by-nc-sa/4.0/ /dk/atira/pure/core/document/licenses/cc_by_nc_sa</small>	Add relation... Edit ↓ ↑ —
CC BY-SA <small>http://creativecommons.org/licenses/by-sa/4.0/ /dk/atira/pure/core/document/licenses/cc_by_sa</small>	Add relation... Edit ↓ ↑ —
CCO <small>https://creativecommons.org/publicdomain/zero/1.0/ /dk/atira/pure/core/document/licenses/cc_public_domain</small>	Add relation... Edit ↓ ↑ —
GNU GPL <small>http://www.gnu.org/licenses/gpl.html /dk/atira/pure/core/document/licenses/gnu_gpl</small>	Add relation... Edit ↓ ↑ —
GNU LGPL <small>http://www.gnu.org/licenses/lgpl.html /dk/atira/pure/core/document/licenses/gnu_lgpl</small>	Add relation... Edit ↓ ↑ —
Other <small>/dk/atira/pure/core/document/licenses/other</small>	Edit ↓ ↑

These can be edited in the URI field for each classification:



view as: List Table

Edit

URI
/dk/atira/pure/core/document/licenses/cc_by_nd

Term *
CC BY-ND

Description
http://creativecommons.org/licenses/by-nd/4.0/

Cancel Update

CC BY-NC-SA